English-Ukrainian Parallel Corpus of IT Texts: Application in Translation Studies

Khrystyna S. Mandziy [1], Uliana V. Yurlova [1] and Marianna P. Dilai [1]

[1] *Lviv Polytechnic National University, Bandera str. 12, Lviv, 79013, Ukraine*

Abstract

The research is devoted to the application of corpus linguistics technologies in translation studies and practice. In particular, this paper describes the ways in which English-Ukrainian parallel corpus of IT texts can be used to ensure proper translation of IT terms from English into Ukrainian and to identify the main methods of their translation. The corpus is created using Sketch Engine tools. To make parallel concordances easy to use in the process of translating texts in the MS Word text editor, we created a macro program that automatically adds concordance hyperlinks for key English IT terms from the parallel corpus. The code of this macro is written in the Visual Basic programming language. We see the theoretical value of this study in expanding the scope of research on IT terminology, determining and systematizing the peculiarities of its translation from English into Ukrainian. The practical value of the research findings is that the created parallel corpus of IT texts and macro program will be useful for linguists, translators and IT specialists.

Keywords

Parallel corpus, IT texts, IT terminology, translation, Sketch Engine, macro program.

1. Introduction

Today the importance of corpus linguistics is constantly growing. This is due to the practical significance of this field of linguistics and its impact on language study in general. Corpora provide researchers with access to a great variety of natural language data, which becomes the basis of language research at all levels. In addition, they are of great practical importance in lexicography, translation practice, foreign language teaching, machine translation, etc. [8]. Due to the fact that the analysis of corpus data is mainly carried out using an empirical approach, and also because the careful selection of texts for corpora ensures their representativeness, corpus research allows avoiding subjectivity and enables unbiased study of language [19, p. 122]. Therefore, the **topicality** of this paper is determined by the importance of the development of Ukrainian corpus linguistics, in which there is still a large gap in the parallel Ukrainian-foreign corpora construction. The choice of the subject matter of the corpus texts is connected with the rapid growth of the information technologies sphere in Ukraine. IT terminology is changing and evolving rapidly. The translation of such terminology from English into Ukrainian needs additional attention.

The **aim** of the paper is to apply the English-Ukrainian parallel corpus of IT texts in the study of the methods and ways of rendering English IT terminology into Ukrainian and develop a macro program enabling automatic concordance hyperlinks for the key English IT terms from the parallel corpus.

The English-Ukrainian parallel corpus of IT texts was created using Sketch Engine tools. It contains the texts of four programming textbooks in English and their translations in Ukrainian, namely “Java programming for Kids, Parents and Grandparents”, “Python Tutorial”, “Dive into Python 3”, “Learn You a Haskell for Great Good!”, “The C Programming Language” and two web

COLINS-2022: 6th International Conference on Computational Linguistics and Intelligent Systems, May 12–13, 2022, Gliwice, Poland
EMAIL: khrystyna.mandziy@gmail.com (K. Mandziy); uliana.yurlova.mfpl.2021@lpnu.ua (U. Yurlova); marianna.p.dilai@lpnu.ua (M. Dilai)
ORCID: 0000-0002-1216-9538 (K. Mandziy); 0000-0002-8577-0404 (U. Yurlova); 0000-0001-5182-9220 (M. Dilai)

sites with technical documentation for the Bootstrap framework and the php scripting language. The total corpus volume is above 300 thousand words. The procedure of corpus compiling is described in [14].

The research is based on the theoretical provisions of corpus linguistics concerning application of parallel corpora in translation studies and practice, developed by scholars such as J. Sinclair, M. Baker, K. Aijmer, D. Biber, W. Teubert, O. Demska-Kulchytska, M. Shvedova and others.

The corpus linguistics techniques, including keywords, collocation and concordance analyses, are used to extract a list of the key IT terms in the corpus and determine their linguistic properties. Furthermore, a contrastive analysis of the extracted English IT terms and their Ukrainian translations is carried out focusing their common and divergent semantic features, metaphorisity, structural types.

The macro program that identifies IT terms in MS Word texts and adds hyperlinks to the corresponding concordances in Sketch Engine was written in the Visual Basic programming language.

We see the theoretical value of the study in expanding the scope of research on IT terminology, defining and systematizing the peculiarities of its translation from English into Ukrainian. The practical value of the research is that the parallel corpus of IT texts and the macro program will be useful for linguists, translators and specialists in the field of information technology.

2. Theoretical Framework

2.1. Parallel corpora

According to Ana Frankenberg-Garcia, a parallel corpus is “a computerized collection of texts in one language aligned with their translations into another language”. Parallel corpus can provide automatic access to various features of translated texts that up to now have not been possible to study in a systematic way [1, p. 142]. Parallel corpora are not limited to two languages, but may contain several languages of translation. Scholars also differentiate between unidirectional (e. g. translation only from English into Ukrainian), bidirectional (translation in both directions) and multidirectional parallel corpora (translation into several languages simultaneously). The essential step in parallel corpus creation is to align the source texts and their translations, annotating the correspondence between the two at the sentence or word level. This can be accomplished using a computer program or by manual analysis. It should be noted that the automatic alignment of parallel corpora is not a trivial task for some language pairs [25, pp. 21-22].

Parallel corpora are widely in applied linguistics and are a powerful tool in language teaching and learning, translation studies and translator training, machine translation, contrastive and comparative linguistics, terminology studies and lexicography [12].

In recent decades, Ukrainian corpus linguistics has developed quite successfully and there are already several corpora of the Ukrainian language. One of the greatest achievements of the Ukrainian corpus linguist is the Ukrainian National Linguistic Corpus [26], which contains more than 100 million word usages. It is developed by the Ukrainian Language and Information Fund of the National Academy of Sciences of Ukraine under the supervision of Academician of the National Academy of Sciences of Ukraine V.A. Shyrokov [27, p. 103]. The corpus is included in the state register of scientific objects that constitute national heritage. The corpus consists of works written in the modern Ukrainian language, and the aim of the compilers is to collect all the materials created during its existence and development (for about two hundred years). Corpus texts are not limited to genres or styles, users can select the appropriate subcorpora and filter their results [30, p. 48].

The General Regionally Annotated Corpus of Ukrainian (GRAC) is a collection of more than 90,000 texts of various genres created during the entire history of modern Ukrainian literature (since 1816) [10]. GRAC contains a representative sample of texts, on the basis of which researchers have the opportunity to search, process results, form their own subcorpora and concordances. The corpus contains regional and morphological markings. The main contributors of the corpus are Maria Shvedova, Ruprecht von Waldenfels and Vasyl Starko.

The Institute of Philology of the Taras Shevchenko National University of Kyiv has developed another very valuable project, namely The Ukrainian Text Corpus (KTUM) [15]. It is built in the form

of information and reference system. This corpus contains over 3 million word forms and serves to solve various linguistic problems [30, p. 51].

Although Ukrainian corpus linguistics already has a lot of valuable achievements, some of its areas need more attention. In particular, it concerns the creation of parallel corpora, in which texts in Ukrainian are aligned with texts in other languages. It is worth mentioning the parallel corpora that are already publicly available on the Internet: Parallel Ukrainian-Russian and Russian-Ukrainian subcorpus of the Russian National Corpus [24] and the Polish-Ukrainian Parallel Corpus [23].

Thus, one of the promising and relevant areas within corpus linguistics in Ukraine is the development of parallel corpora, which could be widely used by linguists, translators, specialists in artificial intelligence and machine translation, as well as language learners.

2.2. The concept of a term and research on IT terms

The concept of a term as an object of many disciplines does not have one universally accepted definition. A term can be defined as "a unit of lexical level (a word or a collocation) that denominates some concept of respective domain of human endeavour and forms functional thematic class of the field vocabulary and is a natural element of the terminology fund" [18, p. 18]. It is characterized by such properties as: unambiguity, accurate transmission of the concept, existence exclusively within a specific terminology system, a clear definition that explains the meaning of the term [29, pp. 76-88].

Terminology in the field of information technology began its formation only in the middle of the last century with the development and spread of computers. The first researchers of the terminology of this field were O. Reformatskyi, H. Vynokur, and its theoretical provisions were further developed by a number of modern linguists, including O. Superanska, V. Danylenko, L. Buianova, T. Panko, S. Leichik and others. A significant contribution to the translation studies of the IT terms was made by Ukrainian linguists I. Korunets, V. Karaban, T. Kyiak.

Adequate translation of IT terminology is impossible without a full translator's understanding of terms in the source language. The specificity of IT terminology lies in its heterogeneity associated with the simultaneous existence of unambiguous well-known terms and ambiguous vague terms that constantly appear in IT discourse. In many cases, the context is crucial for deciding on the translation of a particular term in a particular situation.

The importance of the study of IT terminology is determined by the necessity of its elaboration and standardization. Terminological standards are often missing when creating new IT terms, so they do not always meet all the requirements for a comprehensive definition. Other problems related to this issue include incomplete definitions, insufficient representation of relations in the system of terms in defining the concept, use of general literary words instead of generic terms [20].

Further research in IT terminology is the key to improving the practice of IT text translation.

3. Methodology

An important stage of research preparation is the choice of methodology, i. e. a certain system of actions and a set of approaches to solving problems. The methodology determines what will be the specifics of scientific activity, how it will be organized and analyzed.

The database for the English-Ukrainian parallel corpus consists of the texts of five programming textbooks and two sites with technical documentation (total - more than 300 thousand words). These texts meet the established selection criteria, namely they relate to the field of information technology (programming); are available in the original in English and translated into Ukrainian; are publicly available in electronic form on the Internet. The first textbook deals with the Java programming language: "Java programming for children, parents, grandparents" by Yakov Fain (2004) translated by Vitalii Kostiuchenko (2014) [31]. There are also three textbooks on Python programming: "Python Tutorial" by Guido van Rossum and Fred L. Drake Jr. (first edition 2001) translated by Serhii Kuzmenko (2005) [9], "Dive into Python 3" by Mark Pilgrim (first edition 2001), translator is not mentioned (2020) [17]; "The C Programming Language" by Brian W. Kernighan and Dennis M. Ritchie (1988) translated by Vitalii Tsybuliak [4]. The fourth textbook – "Learn You a Haskell for

Great Good!" Mirana Lipovaca (2011) translated by Anna Leliv, Semen Tryhubenko, Bohdan Penkovskyi, Maryna Strelchuk and Tetiana Bohdan (2017) [16].

The technical documentation texts that have been added to the corpus concern the Bootstrap framework [5] and the PHP scripting language [22]. These texts were translated by programmers Konstiantyn Tretiak and Mykola Pukhalskyi.

The procedure of the parallel corpus creation using corpus manager and text analyzer Sketch Engine as described in [14] included some technical standardization and formatting of texts, and aligning sentences with their translation (in .xlsx format). It should be noted that the corpus contains more than 10 thousand aligned sentences. Sketch Engine automatically adds morphological markup to downloaded English texts. During this process, each word form is assigned a POS tag ("Part of Speech"), which contains information about belonging to a part of speech and about some grammatical categories (cases, tenses). The accuracy of the results of automatic tagging is 98% and usually errors occur only when using words in their uncharacteristic meanings [2, p. 28]. Automatic tagging of Ukrainian texts is not available in Sketch Engine.

Figure 1 presents general quantitative information about the English and Ukrainian parts of the created parallel corpus. For both languages, Sketch Engine automatically calculates the number of tokens, words (and unique word forms), sentences, paragraphs and uploaded documents. Automatic assignment of certain attributes ("word", "lemma", etc.) to elements of English text significantly affects the usability of the corpus, as it allows you to search for several word forms.

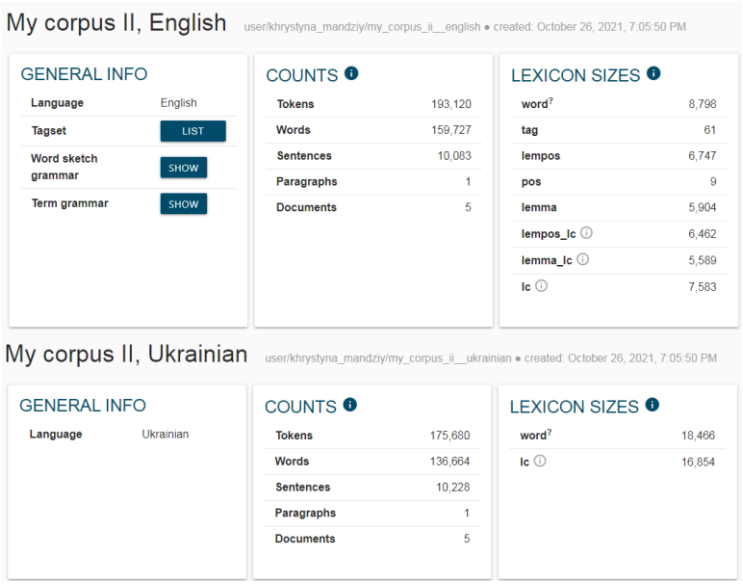

Figure 1: General information about the English and Ukrainian parts of the parallel corpus

The next step was to retrieve key IT terms from the created parallel corpus. The analysis of keywords and phrases is a statistical method used to study the salient words in the corpus. Keywords are the words that are more frequent in the corpus in comparison with the "reference corpus". It is believed that this method reduces the bias of researchers in content analysis and allows to objectively identify keywords and phrases for further analysis [21, p. 33]. We will use this method of analysis to determine the key terminology from the texts of our corpus. Sketch Engine makes it possible to extract the list of keywords (single-words and multi-word terms) automatically using keywords function (see Figure 2). The reference corpus in our case is British National Corpus (BNC) [6]. For the further linguistic and translation analysis we used the top 205 English terms extracted from the parallel corpus. The most frequent of them are the following: *function* (1458), *variable* (763), *file* (704), *subclass* (642), *module* (450), etc.

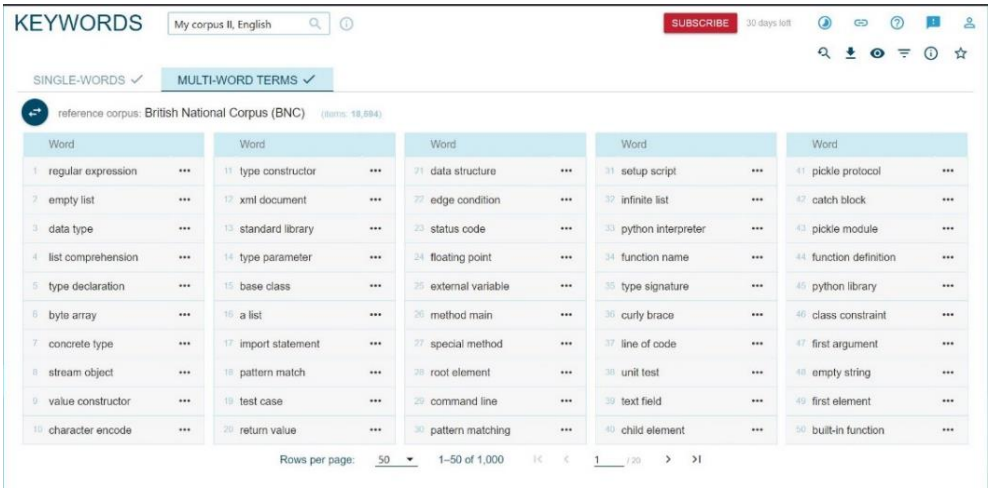

Figure 2: Keywords **(**multi-word terms) extracted from the English texts of the parallel corpus

Parallel concordance analysis allows to analyze translations of texts, compare languages at different levels. Another method for translation analysis is contextual analysis, which is also based on concordances construction. Contextual analysis provides a better understanding of when and why language units are used.

The last step was to find Ukrainian equivalents to the key IT terms in English using the parallel concordance function (see Figure 3). This feature made it possible to trace the use of terms in context and to determine the method of translation of each terminological unit found.

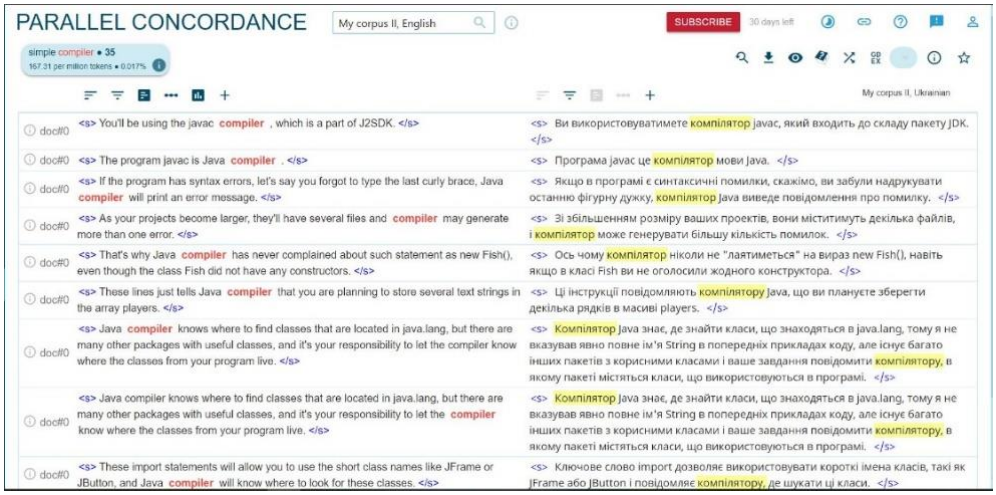

Figure 3: The search result of the Ukrainian equivalent for the term “compiler” in the parallel corpus

Parallel concordance is a convenient way to represent words in their contextual environment in two languages. It allows you to quickly get information about the use of words and ways to translate them. However, when working with large volumes of texts containing unknown IT terminology, finding each individual word can be time consuming. To solve this problem, we created a macro program that will look for parallel concordances in the Sketch Engine containing IT terms from user texts. This will greatly simplify access to the corpus data.

One of the most popular text editors that can be used for both personal and work or study purposes is the MS Word processor. A significant advantage of this software for linguists is the ability to create programs for processing natural language texts based on object-oriented programming language Visual Basic in the embedded environment Visual Basic for Applications (VBA). VBA is a programming tool designed to build macro programs. Macro is a sequence of characters whose input results in a different sequence of characters that performs specific predefined tasks. Macro commands are entered into the VBA editor module.

We developed a macro for text analysis, which will allow the user to quickly use the database of English-Ukrainian parallel corpus of IT texts to translate their own texts in the MS Word editor. The macro searches the text for words given in the glossary of the key IT terms and add hyperlinks to the corresponding concordance in Sketch Engine. Figure 4 shows a model of the process of using such a macro by the user. The model was created using the system of notation BPMN ("Business Process Model and Notation").

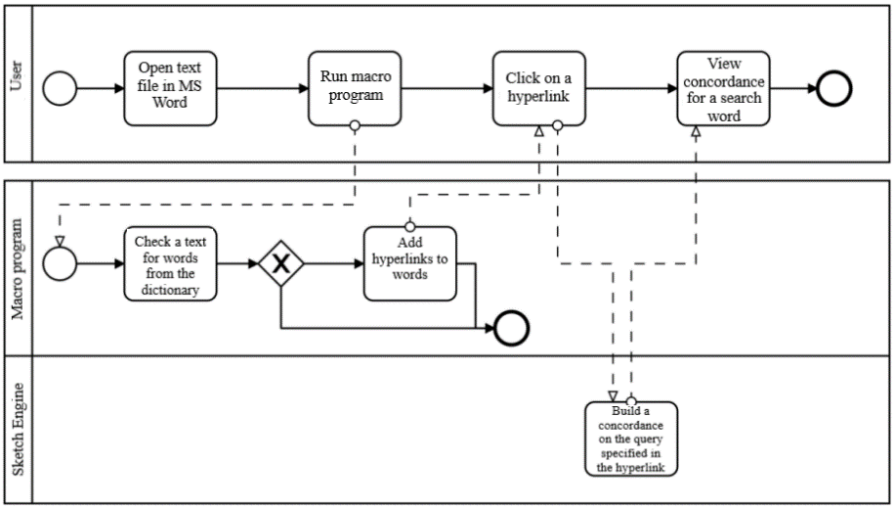

Figure 4: The process of using a macro to search for words and add hyperlinks to the text

The model consists of three tracks, each of which is responsible for a separate part of the process – user actions, the sequence of macro actions and Sketch Engine actions caused by the macro. The first step of the process is performed by the user, opening the program MS Word and running the macro. To start the macro, the user should select a file with English text. The first step of the macro is to check the text for words from our glossary of the key IT terms. If the corresponding words are found, they are accompanied by a hyperlink to the concordance in Sketch Engine, built on the basis of texts from our parallel corpus. If no word in the text belongs to the dictionary, the macro does not take any action. The user can open one of the added hyperlinks by simultaneously holding down the CTRL key and left-clicking on that hyperlink. The user will then be redirected to a new browser window with a concordance for the selected word in the Sketch Engine. The user profile in Sketch Engine needs to have access to our parallel corpus. This access is provided individually.

Figure 5 shows a VBA window with a macro code. It consists of two procedures: readWords () and createHyperlinks ().

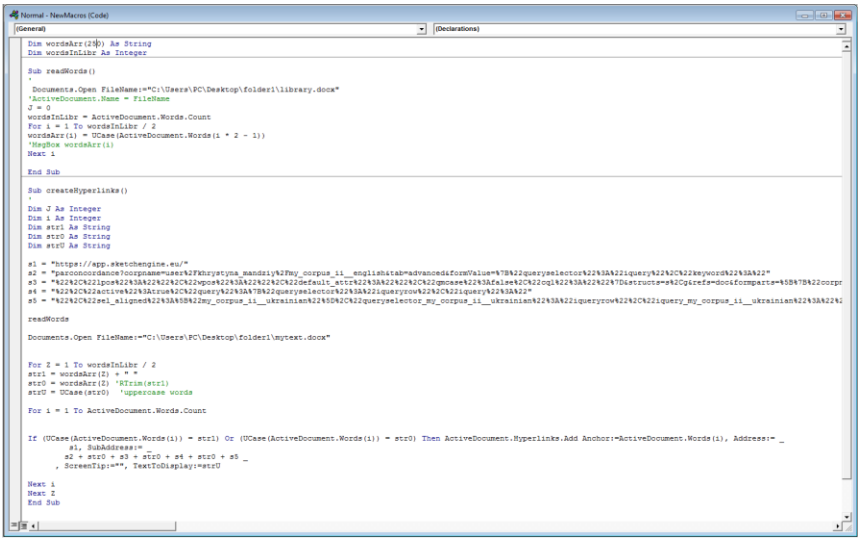

Figure 5: Macro program code

The readWords () procedure opens the specified file (FileName), which contains a glossary of IT terms, reads all the words in the file, and adds them to the wordsArr array. Execution of the createHyperlinks () procedure begins with the user selecting a file with the text to be processed (opened using the BrowseForFile () function). In this text, each word is checked for compliance with words from the dictionary (capital letters are ignored). If the word is in the dictionary, it is assigned a hyperlink:

If (UCase (ActiveDocument.Words (i)) = str1)
Or (UCase (ActiveDocument.Words (i)) = str0)
Then ActiveDocument.Hyperlinks.Add Anchor: = ActiveDocument.Words (i),
Address: = _s1, SubAddress: = _s2 + str0 + s3 + str0 + s4 + str0 + s5_

The hyperlink consists of 6 parts, where str0 is a search word, s1 is a link to the Sketch Engine program, s2 is a link to the subcorpus concordance, s3, s4, s5 is a combination of search filters in the English and Ukrainian parts of the corpus.

Figure 6 presents the text processed by the macro. 10 words from the text were found in the glossary of the key IT terms. Each term has a hyperlink which leads to a concordance of the corresponding term.

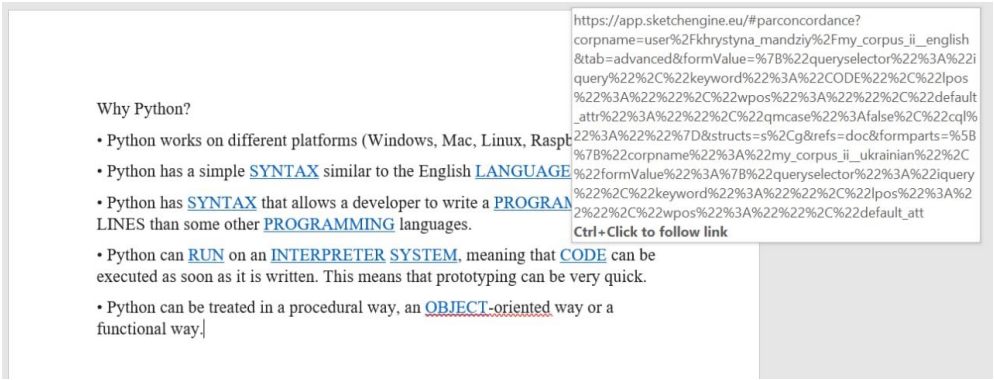

Figure 6: Text in MS Word with hyperlinks to concordances in Sketch Engine

Figure 7 shows the concordance constructed for the word “syntax”. The user is redirected to this concordance when clicking on the hyperlink added to the text in MS Word.

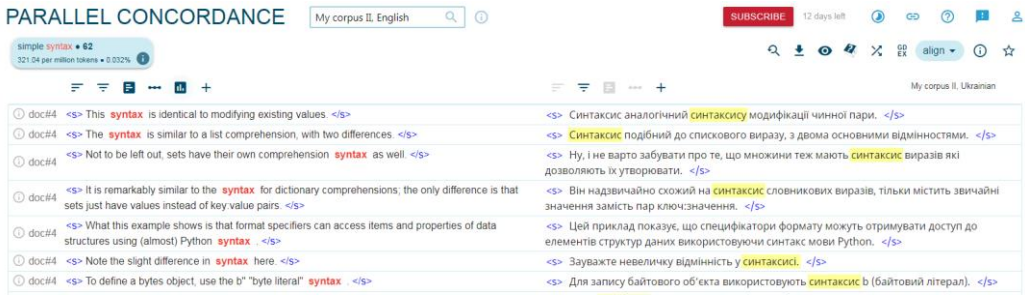

Figure 7: Parallel concordance for the word “syntax”

The created macro gives access to parallel concordances for words directly from the MS Word editor. This significantly saves time by working with the text. By opening the concordance in Sketch Engine, the user can quickly get acquainted with the context of terms usage and possible translation options.

3. Results and discussion

The aim of the study at this stage was to analyze 205 IT terms in English and their translations into Ukrainian in terms of type and method of translation. We have established that the most typical for

the translation of IT terminology are the following methods: transliteration, calque (loan translation), explication and equivalent method.

The most common translation method used in this investigation is calque or loan translation that is a translation technique by which a word taken from one language and translated in a literal or word for word way to be used in another [7]. As a result, we found out that 84 out of 205 terms are translated using this method, for example: *technology – технологія, invent – винаходити, pointer – покажчики, complicated declarations – складні оголошення, machine-readable form – машинопрочитний вигляд*.

A statistically insignificant number of IT terms are translated using transliteration, that is a mapping from one system of writing into another, word by word, or ideally letter by letter. As a result of transliteration method a reader should be able to reconstruct the original spelling of unknown transliterated words. To achieve this objective, transliteration may define complex conventions for dealing with letters in a source text which do not correspond with letters in a target text. This method is used when there are no means of verbalization in the target language for the given word, so the term is "taken" from the source language though it is written with the means of a target one. The results showed that 25 terms out of 205 were translated using transliteration, for example: *hypertext – гіпертекст, computer – комп'ютер, server – сервер.*

Translation equivalence is the similarity between a word (or expression) in one language and its translation in another. This similarity results from overlapping ranges of reference. A translation equivalent is a corresponding word or expression in another language. As a result, we found out that 71 out of 205 terms are translated using this method, for example: *mode – режим, toggle – перемикач, button – кнопка, tooltip – підказка, whitespace – пропуск*.

The characteristics of explication as a means of adequate translation of non-equivalent vocabulary have been analyzed. In scientific and technical texts we can come across specialized terminology that in translation has no direct equivalents, and therefore, the main task of the translator of technical literature is a pragmatic adaptation of the original text preserving its form and content. As a result of the research, we found out that 13 out of 205 terms had been translated using this method, for example: *the C Programming Language – мова програмування С, UNIX – operating system операційна системи Unix, ASCII character set – набір знаків ASCII.*

The methods of reduction and specification were also identified, but they are used with less frequency. In total, we encountered only 10 cases of specification and 2 case of reduction, for example: *isolated fragments – окремі фрагменти коду, newline – знак нового рядка, execution – виконання програми, source program – вихідний текст програми, header names – назви файлів заголовка, interchange sorts – взаємозамінних алгоритмів сортування.*

It should be noted that during the analysis of IT term translation methods, we identified a number of units that were translate using different methods. For instance: *personal computers – особисті комп'ютери (calque), комп'ютери (reduction); machine – машина (calque), прилад (equivalence); compiler – компілятор (calque), програма-компілятор (specification); calculator – калькулятор (transliteration), обчислювач (equivalence).*

Thus, based on the results obtained we can conclude that the method of calque is the most typical. Methods of transliteration and explication are used much less often, but they are applied when it is difficult to convey the meaning of a term using the lexical means of the target language. Translation methods such as specification and reduction can be used when translating terms in the field of information technology, but translators rarely use them in their practice.

The analysis of lexical-semantic groups of 166 noun terms (other parts of speech are not taken into account in this analysis) was carried out on the basis of the classification of I. Mentynskaya [13, pp. 28-29]: names of knowledge areas, e.g. *programming* and *functional programming*; names of specialists, e.g. *programmer*; names of units of information, e.g. *byte*; software names and their elements: types of application software (*constructor, installer, text editor*) and software elements (*command line, plugin*); terms related to the processes of working with information: collection, storage, processing and transmission, e.g. *initialization, caching, import*; terms related to the Internet and Internet communication, e.g. *browser, cache, web site, web service*.

The groups of terms described above make up only 27% of the studied noun terms. The rest of the terms cannot be attributed to any of the selected groups. Therefore, having analyzed them, we distinguish three more groups: terms of programming theory, names of graphical interface elements

and names of graphic symbols. The most numerous is the group of terms of programming theory (88 terminological units): program components (*class, expression*), values (*argument, variable*), methods in programming (*declaration, permutation, concatenation*). Terms denoting elements of the graphical interface include *button, carousel, checkbox*. The last is a group of graphic symbols. These characters are part of the software code, or are used by users to interact with the software, e.g. *backslash, hash*.

Figure 8 shows a pie chart illustrating the quantitative ratio of lexical-semantic groups of the key IT terms in the English-Ukrainian parallel corpus of IT texts.

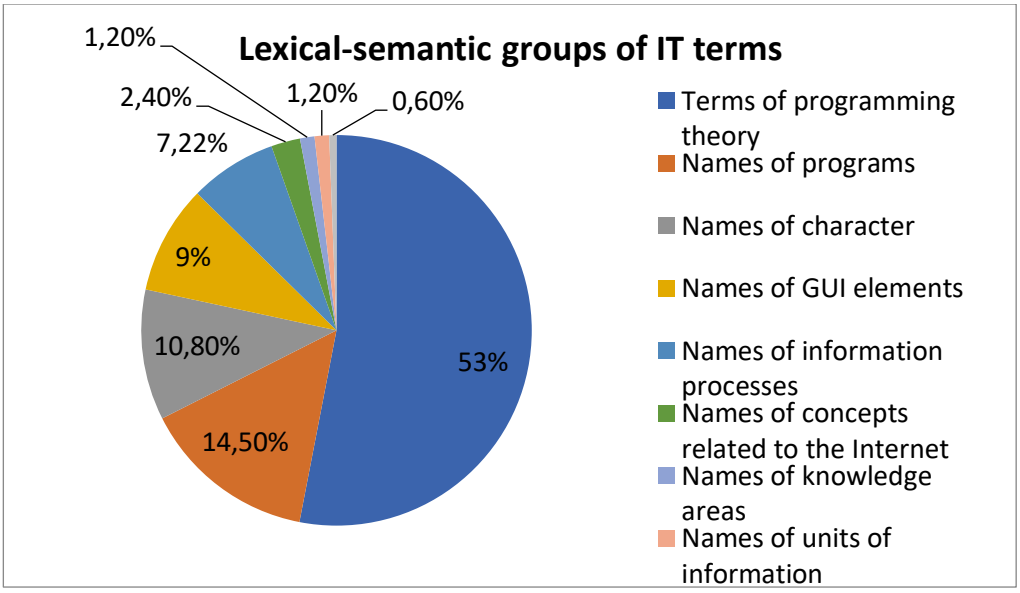

Figure 8: Quantitative ratio of lexical-semantic groups of IT terms

Thus, the most numerous lexical-semantic groups of IT terms in our corpus are the terms of programming theory and software names and their elements. This is due to the specifics of the texts that make up the corpus – they relate to programming and software development.

The next stage of our study was the analysis metaphorical IT terminology in English and Ukrainian. The researched material contains 166 simple English terms and their Ukrainian equivalents, among which 85 English (51%) and 72 Ukrainian (43%) metaphorical terms were identified. The sample analysis shows that IT terms are often formed by metaphoric rethinking commonly used words and phrases (for example, *key – ключ, consistency – узгодженість, shell – оболонка*).

Expanding the classification of metaphor domains introduced by V. Celiešienė, S. Juzelėnienė [28, pp. 92-97], we distinguish the following source domains: clothing items, household items, tools, places, people, nature, actions and states, science, social life, position in space. In 18% of cases, English metaphorical terms have been translated into Ukrainian using the method of transcription, so we do not consider terms in the language of translation to be metaphorical, e.g. *cache (*warehouse – the domain of place) – *кеш*, *stack* (rick – the domain of place) – *стек*. In 14% of cases, we trace the change of the source domain when translating metaphorical terms: *padding* (lining – the domain of clothing items) – *відступ* (the domain of position in space), *toggle* (lever – the domain of tools) – *перемикач* (the domain of household items). These terms are translated into Ukrainian by equivalents.

Table 1 presents the results of quantitative analysis of source domains for the metaphorization of IT terms in English and Ukrainian.

Table 1

Frequency of source domains

Domain	Relative frequency (English), %	Relative frequency (English), %	Examples (English + meaning in Ukrainian)	Examples (Ukrainian term + English term)

items of clothing, fabrics	3,5	1	*button (кнопка)*	*стрічка (string)*
household items	14	17,3	*folder (папка)*	*пакет (package)*
tools	8,2	6,4	*brush (пензлик)*	*сітка (grid)*
places	2,4	–	*cache (склад)*	–
people	9,4	17,3	*constructor (конструктор)*	*оператор (operator)*
nature	2,4	2	*tree (дерево)*	*галочка / пташка (checkbox)*
actions and states	28,2	24,8	*call (викликати)*	*пошук (lookup)*
concepts of science	14,2	12,4	*method (метод)*	*код (class)*
phenomena of social life	11,8	11,3	*declaration (оголошення)*	*документація (documentation)*
position in space	5,9	7,5	*redirect (пере направити)*	*навігаційна панель (navbar)*

We distinguish the following groups of features on the basis of which the metaphorical transfer was carried out:1) by form: *галочка / пташка – checkbox*; 2) by function: *folder – папка*; 3) by the mechanism of action: *space – пробіл*; 4) by the nature of the action: *call – викликати*; 5) by size: *snippet – шматок коду*.

Comparing English metaphorical IT terms with their Ukrainian equivalents, we concluded that in most cases (68%) the terms retained their metaphorical meaning in both languages. Since the terminology of computer science was formed on the basis of the English language, Ukrainian terms, even translated by their actual Ukrainian equivalents, usually borrow the metaphorical meaning of English terms.

Terms are divided by structural types into simple, compound and complex. Simple terms consist of one word. They are non-derivative (origin is not motivated in other words) and derivatives (have a motivational basis). For example, *tuple (кортеж)* is a non-derivative term, and *encoding (кодування)* is a derivative because it has a motivational basis *code* and word-forming affixes *en-*, *-ing*. Compound terms have several bases (for example, *navbar (навігаційна панель*)), *dropdown* (*спадне меню*)). Complex terms are combinations of several words: *e.g., test case (тест), web service (веб-сервіс)* [29, p. 19]. Not in all cases the structural types of terms in English and Ukrainian coincide.

Simple terms make up the majority of the studied list of the English IT terms - 70%. Among them, one third are non-derivative terms, and two thirds are derivatives. They are often translated into Ukrainian by transcription (*file – файл*) and equivalent, which is also a simple non-derivative term (*mode - режим*, *tuple – кортеж)*. There are also translations from non-derivative terms to derivatives (for example, *toggle - перемикач*, *clause - конструкція*) and compound (*integer – ціле число*). Simple derivative terms are the most numerous in our sample – more than half of all terms. The most productive affixes are: suffixes -or (*validator*), -er (*interpreter*), -tion (*concatenation*), -увач (*накопичувач*), -ок (*зв'язок*) for nouns, -ate (*instantiate*) for verbs, -able / -ible (*collapsible*), -al (*positional*), -ан-(-ян-) (*складений*), -ов- (*десятковий*), -ев- (-єв-) (*булевий*) for adjectives; prefixes re- (*recursion, redirect*), de- (*debug*), en- (*encoding*), де- (*декодувати*). During translation, in most cases, English derivative terms are translated by derivative terms in Ukrainian (for example, *padding - відступ*). In some cases (5%) a compound term is used in Ukrainian (*snippet – шматок коду, pagination – посторінкова навігація/посторінковий поділ*).

Most of the English compound terms (40%) have the structure "noun + noun" (for example, *placeholder, tooltip, navbar*), followed by the model "adjective + noun" (for example, *whitespace, lowercase*) (37%) and "verb + adverb" (for example, *popover, dropdown*) (23%). Most of these terms are translated into Ukrainian by changing the structural model. Among Ukrainian IT terms there are

only 4 composite derivatives: *metadata - метадані/метаінформація, navbar - навпанель, singleton - одноелементний, traceback - трейсбек*. In other cases, they are usually translated by complex terms (for example, *backslash – зворотній слеш, lowercase - нижній регістр*), simple derivative terms (*checkbox - галочка, tooltip - підказка*) or juxtapositional terms (*placeholder – назва-заповнювач*).

A complex term consists of several elements that are connected in some way. Among them, the most productive are two-component models. There are more Ukrainian complex IT terms in our sample than English ones. In 10% of cases of translation from English into Ukrainian we can see complex terms of other structural types in the original. In 61% of cases, this is a transformation from a compound term to a complex one (for example, *backslash – зворотній слеш, semicolon – крапка з комою*). In other cases, we notice the transformation from simple to complex (*snippet – шматок коду, pagination – посторінкова навігація*).

The findings show that the most productive models of transformations of the structure of terms are the model simple derivative → simple derivative (45.5%), compound → compound (15%) and simple non-derivative → simple non-derivative (14%). We can conclude that when translating IT terms from English to Ukrainian, in 77.5% of cases the structure of the term remains unchanged.

The part-of-speech analysis of terms made it possible to establish the relationship between parts of speech in the terminological system. We have revealed a tendency in IT terminology to nomination (it is dominated by nouns, 80%). The prevalence of nouns in IT terminology is explained by the fact that terms mainly perform a nominative defining function. The results show that nouns are followed by verbs (11%), adjectives (8%) and adverbs (1.5%). There cases of conversion in English, e.g. *indent – відступ/відступати*, *input – вхідні дані/вхідний*. In general, 93% of simple terms retain their part-of-speech affiliation during translation.

4. Conclusions

In many scientific, technological or political fields there is a lack of terminological lexicographic resources which causes problems to translators and results in inconsistent translations. Parallel corpora can be used as a resource for automatic extraction of terms and terminological collocations.

In the course of our research using the data and features of the created English-Ukrainian parallel corpus of IT texts we made an attempt to single out the most common methods of IT terms translation and their linguistic peculiarities. The corpus manager Sketch Engine was used to provide convenient and fast access to the corpus. Furthermore, a macro program was created for the translation of IT texts in the text editor MS Word. The macro program allows you to determine which IT terms from the user's text are used in the corpus and add links to the relevant concordances of these terms in Sketch Engine. The key English IT terms were obtained using the method of keywords and were analyzed in terms of lexical-semantic groups they belong to, as well as sources of metaphorization, structural types and part-of-speech affiliation.

The semantic analysis of the IT terms showed that they belong to 9 lexical-semantic groups: terms of programming theory, program names, symbol names, graphical interface elements names, information process names, Internet concepts, names of knowledge areas, names of information units, names of specialists. In addition, ten source domains of metaphorical terms were identified: clothing items, everyday items, tools, places, persons, nature, actions and states, science, social life, position in space. Concerning the structure of IT terms, the most numerous group in both languages is made up of simple derivative terms. In 77.5% of cases, the structure of terms during translation remains unchanged. Part-of-speech analysis of the terms has proved their nominative nature. The morphological structure of English terms is preserved in about half of the cases during translation into Ukrainian. In addition, the research finding show that the most typical method of translating English IT terms into Ukrainian is the method of calque.

We see the practical significance of the results in the fact that the parallel corpus of texts in the field of information technology is a useful tool for translators, as well as those who study languages or are interested in IT. Concordances built on search queries allow you to quickly learn the meanings of words in different contexts. In addition, the proposed macro program will allow you to access

concordances directly from the text editor MS Word, which will significantly reduce the time spent searching in the corpus.

The theoretical value of the work lies in the contribution to the development of the studies related to the systematization and translation of terminology in the IT field. Systematization allows to identify general trends and predetermine approaches to the study, understanding and translation of IT terms.

The paper outlines the advantages of using a parallel corpus during translation and encourages further improvement of the parallel corpus in accordance with certain translation needs in order to achieve the highest level of equivalence of the translated text and the original text to meet the expectations of the translation recipient. The study opens up prospects for further research on the parallel corpus in order to identify new useful means of facilitating the work of translators. We see prospects in the further expansion of the text base of the created English-Ukrainian parallel corpus of IT texts. This will contribute to its greater representativeness and further linguistic and translation study of English and Ukrainian IT terminology.

5. References

[1] A. Frankenberg-Garcia, Using a Parallel Corpus in Translation Practice and Research, (2012).

[2] A. Kilgarriff, V. Baisa, J. Bušt, M. Jakubíček, V. Kovář, J. Michelfeit, P. Rychlý, V. Suchomel, The Sketch Engine: ten years on. Lexicography: Journal of ASIALEX, 2014.

[3] A. Ye. Konverskyi, Osnovy metodolohii ta orhanizatsii naukovykh doslidzhen [Fundamentals of methodology and organization of scientific research], Tsentr uchbovoi literatury [Center for Educational Literature], Kyiv, 2010.

[4] B. W. Kerninghan, D. M. Ritchie, The C Programming Language. URL: http://programming.in.ua/programming/c-language/227-book-programming-c-kernighan.html.

[5] Bootstrap. URL: Bootstrap 4.

[6] British National Corpus. URL: https://www.english-corpora.org/bnc.

[7] Cambridge Dictionary. URL: https://dictionary.cambridge.org.

[8] D. Biber, S. Conrad, R. Reppen, Corpus linguistics: Investigating language structure and use, Cambridge University Press, New York, NY, 1998.

[9] G. van Rossum, Fred L. Drake Jr., Python Tutorial, 2001. URL: http://docs.linux.org.ua/Програмування/Python/Підручник_мови_Python/.

[10] General Regionally Annotated Corpus of Ukrainian. URL: http://uacorpus.org.

[11] H. H. Lukianets, Osnovni napriamky suchasnykh korpusnykh doslidzhen movy ta perspektyvy yikh podalshoho rozvytku [The main directions of modern corpus studies of language and prospects for their further development], Naukovi pratsi natsionalnoho universytetu kharchovykh tekhnolohii [Scientific works of the National University of Food Technologies] (2012) 127–133.

[12] H. M. Alotaibi, Arabic-English Parallel Corpus: A New Resource for Translation Training and Language Teaching, Arab World English Journal (2017) 319-337.

[13] I. B. Mentynskaya, Tematychna ta leksyko-semantychna klasyfikatsiia ukrainskykh kompiuternykh terminiv [Thematic and lexical-semantic classification of Ukrainian computer terms], Vcheni zapysky TNU imeni V. I. Vernadskoho. Filolohiia. Sotsialni komunikatsii [Scientific notes of TNU V. I. Vernadsky Philology. Social communications] (2020) 26-30.

[14] K. Mandziy, M. Dilai, Osoblyvosti pobudovy ta perspectyvy vykorystannia anhliisko-ukrainskoho paralelnoho korpusu IT tekstiv [Specificity of compiling and perspectives of using of English-Ukrainian parallel corpus of IT texts], Problemy humanitarnych nauk: zbirnyk naukovych prats Drohobytskoho derzhavnoho pedahohichnoho universytetu imeni Ivana Franka. Seriia “Filolohiia” [Problems of Humanities. “Philology” Series: a collection of scientific articles of the Drohobych Ivan Franko State Pedagogical University] (2021) 120–127.

[15] KTUM Corpus of Ukrainian texts. URL: http://www.mova.info/corpus.aspx.

[16] M. Lipovaca, Learn You a Haskell for Great Good!, 2011. URL: https://haskell.trygub.com.

[17] M. Pilgrim, Dive into Python, 2001. URL: https://uk.wikibooks.org/wiki/Пориньте_у_Python_3.

[18] M. Vakulenko, Term and terminology: basic approaches, definitions, and investigation methods (Easterm-European perspective), Terminology Science & Research (2014) 13-28.
[19] O. Demska-Kulchytska, Osnovy natsionalnoho korpusu ukrainskoi movy [Fundamentals of the national corpus of the Ukrainian language], Kyiv, 2005.
[20] O. O. Volkova, Strukturni kharakterystyky anhlomovnoi terminolohii u haluzi media-komunikatsii ta osoblyvosti yii perekladu ukrainskoiu movoiu [Structural characteristics of English terminology in the field of media communication and features of its translation into Ukrainian], In Statu Nascendi, Kharkiv, 2010.
[21] P. Pérez-Paredes, Corpus Linguistics for Education. A Guide for Research, Routledge, London, 2020.
[22] PHP documentation. URL: Ukrainian translation of the PHP documentation.
[23] Polish-Ukrainian Parallel Corpus. URL: http://domeczek.pl/~polukr/.
[24] Russian-Ukrainian and Ukrainian-Russian parallel subcorpora of the RNC. URL: http://www.ruscorpora.ru/search-para-uk.html.
[25] T. McEnery, A. Hardie, Corpus linguistics: method, theory and practice, Cambridge University Press, New York, NY, 2012.
[26] Ukrainian National Linguistic Corpus. URL: http://unlc.icybcluster.org.ua/virt_unlc.
[27] V. A. Shyrokov, O. V. Buhakov, T. O. Hriaznukhina, Korpusna linhvistyka [Corpus linguistics], Dovira, Kyiv, 2005.
[28] V. Celiešienė, S. Juzelėnienė, Metaphorical Nomination in IT Terminology in Lithuanian and English Languages, Journal of Language and Cultural Education (2019) 84-102.
[29] V. M. Leichik, Terminovedenie predmet metody struktura [Terminology: subject, methods, structure], Izdatelstvo LKI [LKI publishing house], Moscow, 2007.
[30] V. V. Zhukovska, Vstup do korpusnoi linhvistyky [Introduction to corpus linguistics], Vydavnytstvo zhytomyrskoho derzharnoho universytetu im. Franka [Zhytomyr Ivan Franko State University publishing house], Zhytomyr, 2013.
[31] Y. Fain, Java programming for children, parents, grandparents, 2004. URL: java4kids.